Reports and Technical Reports | Graduate School of Business and Public Policy (NPS-GSBPP)

2005-12-30

A Performance Metric and Goal-setting Procedure in Deadline-oriented Processes

Doerr, Kenneth H.

http://hdl.handle.net/10945/38091

NPS-CDMR-LM-06-005

CENTER FOR DEFENSE MANAGEMENT REFORM

A Performance Metric and Goal-setting Procedure for Deadline-oriented Processes

30 December 2005

by

Dr. Kenneth H. Doerr, Associate Professor,
Graduate School of Business & Public Policy
Naval Postgraduate School

Dr. Kevin Gue, Associate Professor,
Department of Industrial and Systems Engineering
Auburn University

Approved for public release, distribution is unlimited.

Prepared for: Naval Postgraduate School, Monterey, California 93943

The research presented in this report was supported by the Center of Defense Management Reform of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request Defense Policy and Management Research or to become a research sponsor, please contact:

Attn: Dr. Douglas A. Brook
Director
Center for Defense Management Reform
Graduate School of Business & Public Policy
Naval Postgraduate School
555 Dyer Road, Room 320
Monterey, CA 93943-5103

Tel: (831) 656-3487
Fax: (831) 656-2253
e-mail: dabrook@nps.edu

Copies of the Center for Defense Management Reform Research Reports may be printed from our website www.nps.navy.mil/gsbpp/CDMR

Abstract

We define and show how to set goals against a performance metric for order fulfillment operations in which order requests arrive continuously and in which filled orders are shipped at a specific time each day. Managers in such systems must decide not only on a performance goal for the metric, but also which deadline (target shipment) to assign to each order. We use results from goal-setting theory to establish the performance goal, and then illustrate how best to match arriving orders to deadlines based on their arrival times and expected processing times. We use data from a large distribution center to demonstrate that setting these two parameters in the light of motivational research yields quite different values than doing so with an intuitive method. Moreover, a motivational goal leads to better operational performance; that is, correctly setting up the metric causes more customers to receive their orders sooner.

Key words: Deadlines, Goals, Performance Metrics, Motivation, Bootstrapping, Warehousing, Distribution, Work Design

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

This research was supported in part by Research Initiation Program funds at the Naval Postgraduate School.

THIS PAGE INTENTIONALLY LEFT BLANK

About the Authors

Ken Doerr, is an associate professor of operations management in the Graduate School of Business and Public Policy at the Naval Postgraduate School. His research interests are in work design, and the valuation of technology for logistics.

Kevin Gue is an associate professor in the Department of Industrial and Systems Engineering, School of Engineering at Auburn University. His research interests are in facility logistics, warehousing and distribution, with particular interest in cross-docking operations in the retail and transportation industries.

THIS PAGE INTENTIONALLY LEFT BLANK

NPS-CDMR-LM-06-005

CENTER FOR DEFENSE MANAGEMENT REFORM

A Performance Metric and Goal-setting Procedure for Deadline-oriented Processes

30 December 2005

by

Dr. Kenneth H. Doerr, Associate Professor,
Graduate School of Business & Public Policy
Naval Postgraduate School

Dr. Kevin Gue, Associate Professor,
Department of Industrial and Systems Engineering
Auburn University

Prepared for: Naval Postgraduate School, Monterey, California 93943

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

Success in a distribution network is often thought of as getting product to the customer as quickly as possible, subject to some reasonable cost. We propose and investigate a metric, called the Next Scheduled Departure (NSD) that links warehouse performance to customer service. The key insight behind our metric is that picking orders in a warehouse is a continuous process, but the next step in the chain -- transportation -- is a scheduled, batch process. NSD links the two processes in such a way that an increase in NSD increases customer service directly.

The operational problem we investigate is to set a cutoff time and a performance goal against the NSD metric. NSD is a tune-able metric in the sense that managers establish a cut-off time that determines which orders count toward a day's performance. All orders arriving before the cutoff time are targeted to ship on the truck scheduled to depart next. The performance goal is stated as a percentage of orders that actually make the target, and get shipped on the next truck. In general, the earlier the cutoff time, the greater the chance that all orders will make it onto the next truck and so management can set a higher goal. But the earlier the cutoff time, the greater the number of orders that might be ignored because they are not due until the next day, even though they might have been processed immediately upon arrival --- implying that real (vice measured) performance could suffer. Also, management feels proximal goals are more motivational than distal ones, and hence, wants to maintain a later cutoff time, closer to actual truck departure.

We use findings from the goal-setting literature to explore the tradeoff between the cutoff time and the goal, and demonstrate the significance of behavioral research on this operations problem. However, we also encounter and area where there is no empirical guidance for model building: While there is a great deal of guidance in the literature about e.g., the size of the goal setting effect, less is known about the time sensitive nature of this effect. We build a model that incorporates a temporal factor into the goal-setting effect, and use simulation to select robust values for the cutoff time and performance goal in light of that temporal factor.

THIS PAGE INTENTIONALLY LEFT BLANK

Introduction

A typical order fulfillment operation receives orders from customers in a continuous stream and prepares them for shipment at a specific time or times during the day. For example, orders to an internet retailer's distribution center arrive throughout the day, but shipments leave the warehouse only when the package carrier arrives for a pickup near the end of the day. Make-to-order manufacturing systems sometimes operate according to this model. In one case the researchers know of, a custom engraving service receives orders throughout the day and completes as many as possible before the package carrier makes its final pickup in the early evening.

Managers of these systems assess performance with a variety of metrics related to customer service, quality, safety, financial measures, and so on. With respect to operations, they generally strive to achieve a high level of customer satisfaction at the lowest possible cost. Customers usually hope to receive their orders "as soon as possible," and so a common measure of customer service is how long it takes to process an order. Labor cost is a major concern because order fulfillment operations are usually labor-intensive, and so a common measure is orders processed per person-hour.

The intended effect of these metrics is that workers will work as efficiently as possible, and customers will receive their orders as soon as possible. Unfortunately, the system does not always work this way. For example, at a large order fulfillment center in California with whom the researchers have worked, performance was measured by flow time, or the average time from receipt of an order to the time it was ready to ship. The corporate goal was to have this average less than 24 hours. In an effort to meet the goal, managers often held workers late on Friday nights so the average order flow time would be reduced. But, in this case, workers returned to work Monday morning to find the orders they filled Friday night still waiting on the shipping dock because there had been no pickups on the weekend.

It is important to note that this behavior is completely rational. Supervisors and workers were doing what was required to improve performance, as defined and recorded by the reigning flow time metric. Performance evaluations, raises, and job security hung in the balance. Moreover, management should have been pleased to see that the metric actually motivated workers: supervisors and workers were striving to improve performance against it.

Yet, as a metric at this order fulfillment center, flow time induced unwanted behavior because it failed to incorporate the batching phenomenon inherent at some point in nearly all supply chains. Had workers been able to place completed orders directly on a conveyor destined for their customers, flow time would have been an appropriate measure; an improvement in the metric would have resulted in a direct benefit to the customer. In the presence of a discontinuity in the supply chain, where it transitions from a continuous to a batch process, a metric such as flow time does not necessarily represent such a direct correlation.

The Metric

The authors proposed a different metric, called *Next Scheduled Deadline* (NSD), to management. This metric, the authors believe, more precisely captures the goals of an order fulfillment operation because it incorporates directly the batching that occurs in transportation. The metric is recorded like this: Suppose that the package carrier makes its final pickup at 17:00 each day (for clarity, we use 24-hour clock time throughout). Management at the order fulfillment center establishes a cut-off time, say at 13:00 each day, such that any order arriving before 13:00 is due on the 17:00 truck. The metric records the fraction of packages arriving between 13:00 yesterday and 13:00 today that have been shipped by the time the 17:00 truck departs today. For example, if 2,000 orders arrive between 13:00 on Day 1 and 13:00 on Day 2, and 200 orders remain unshipped after the truck leaves at 17:00 on Day 2, performance on the NSD metric is 90%.

Next Scheduled Deadline (NSD) is a more sensible metric for such an order fulfillment process for at least three reasons. First and foremost, for a given cutoff time, an increase in the metric indicates a direct improvement in customer services—some customers receive their packages a day earlier than they otherwise would. This is not true of flow time, where a decrease in the metric may or may not result in customers getting their packages sooner.

Second, the metric promotes sensible workforce scheduling by recognizing the transition from a continuous to a batch process (avoiding the “work late Friday, ship on Monday” problem). Specifically, workers are motivated to work at an accelerated pace when it matters most—just before the deadline. Moreover, workers have a structure within which they can devise task strategies to improve performance in ways other than simply working harder. For example, if there is often a large influx of orders around noon, workers might take early lunches in order to be prepared for the spike in orders.

Third, should the firm decide to publish a cutoff time (e.g., orders arriving before the cutoff are “guaranteed” to arrive the next day), it can be used to manage customer expectations and, therefore, to improve the perception of service. For example, one specific internet-based software provider promises orders will arrive the next day if they are placed before 02:00 the night before. In such a system, the customer ordering at 01:30 expects that his order will arrive the next day (actually the same day, because he orders just after midnight), while the customer ordering at 02:30 has *no* expectation of receiving his package. In this case, the firm manages the expectations of its customers—an ideal situation for a service company. Of course, the cutoff time published to the customer need not, and probably should not be the same as the cutoff time given to employees. Customer "guarantees" regarding a cutoff time should be met with a consideration of shortage costs and service expectations: being at or near 100% is desirable. But as this study will illustrate, cutoff times established to motivate employees need to be more challenging than that: a goal that can be met 100% of the time is not a motivational goal. This paper will primarily deal with the latter concern—setting a motivational goal. This text will return to the issue of cutoff times for customer service and the likely interaction with a specific motivational goal in the results discussion.

Problem Statement

Implementing a motivational goal on the Next Scheduled Deadline metric requires the firm to answer two related questions: what target should be given to workers with respect to the NSD metric, and what should be the cutoff time? It is easy to see that these questions are related: a later cutoff time means fewer packages arriving before the cutoff will make it on the departing truck; therefore, NSD performance will be lower. In the previous example, if the cutoff time had been 15:00 instead of 13:00, management should expect a lower percentage of packages to make it on the truck. Had the cut-off time been 10:00, the percentage would likely be higher.

The cutoff time, then, becomes a sort of knob with which management can adjust performance against the metric. In the absence of motivational effects on workers, such adjusting is simply an exercise in accounting and does not affect true customer service (we assume for now that the firm does not publish the cutoff time to customers). To see this, imagine a stream of orders over several days, all processed and sent to customers via departing trucks. In the absence of motivation, moving the cutoff time has no effect on the completion time of any orders; actual customer service is the same.

When a motivational goal is given on the NSD metric, the cutoff time affects actual customer service; so, it should be set to the time that confers the greatest advantage. The following analysis will demonstrate this point.

The major research question of this study is: *Given an arrival process and processing time distribution, how does a knowledge of the motivational effect of goal setting inform the establishment of (1) the cutoff time and (2) a performance goal when the objective is to minimize the expected time customers must wait for their orders?*

In the next section, this text will review relevant literature in performance measurement, motivation, and goal-setting theory and argue that the work contained herein begins to fill a gap at the boundary of operations management and organizational behavior. Section 3 describes the proposed model, and Section 4 demonstrates it using data from a field site. The study closes with conclusions for the research community and descriptions of how firms might use the following results.

Literature Review

Research on performance measurement is primarily in the domain of managerial accounting. But at least since the publication of *Relevance Lost* (Johnson & Kaplan, 1987), there has been a sharp awareness of the sometimes dysfunctional connection between performance measures and an organization's ability to not only assess, but to meet its tactical and operational objectives. Recent performance measurement research has increasingly focused on non-financial (as well as financial measures), and especially on non-financial measures which help employees understand how their work helps to meet customer needs (Euske & Zander, In press). There is a link between performance measurement and employee behavior. Performance measurement changes organizational performance; often, performance measurement systems encourage behavior that does not advance operational objectives (Kerr, 1975).

One of the primary gaps between the operations perspective and the financial perspective has to do with the impact and use of time (Neely & Austin, 2002). Not just the need for prospective—or at least real-time measures of performance versus the traditionally backward-looking historical financial measures—but also the proper accounting for the utilization of time itself, as time is connected both with corporate strategy and customer satisfaction (Stalk & Hout, 1990). Managers increasingly incorporate operational metrics to insure that performance measurement encourages the right behavior. Many of these metrics, especially those associated with customer service, are deadline-oriented, e.g., "critical team objectives like filling an order within 24 hours" (Meyer, 1998).

One of the key factors in changing behavior is motivation, and there is an increasing awareness of this link between motivation and performance measurement. In writing on the emerging trends in performance measurement, Austin and Larkey (2002) assert, "Motivational measurement is explicitly intended to affect the people who are being measured. […] Used in this way, measurement is an attempt to control individual activity which, it is assumed, will not be congruent with organizational objectives, absent the measurement" (p. 337).

Motivation, of course, is not the domain of management accounting but of organizational behavior (OB). In the literature on motivation, goal-setting theory is often given a central place (Mitchell & Daniels, 2001). Goal-setting theory has also been pointed to as an important area in which OB research can inform operational models (Boudreau, Hopp et al., 2003). In goal-setting theory, the effect of a goal on performance is *mediated* by motivation. That is, setting a goal has little or no effect on performance except *through* the impact it has on motivation. The mechanisms of this mediation have been the subject of considerable research and are fairly well understood (Locke, Latham et al., 1990). Goals, when accepted (i.e., internalized), improve performance through behaviors such as attention, effort, persistence and improved task strategies. Goals are said to direct workers' attention to more productive (in terms of the goal) elements of a task, to increase the effort a worker exerts to accomplish a task and to encourage workers' persistence (to the point of achieving the goal) at the task. Goals also encourage innovation by encouraging the development of improved strategies for accomplishing a task. Thus, there are a number of ways that goal setting, through motivation, can increase performance.

The basic finding that specific, difficult goals have a positive impact on performance has been validated many dozens of times through original work and several meta-analyses (Johnson, Maruyama et al., 1981; Tubbs, 1986; Wood, Locke et al., 1987). The goal-setting effect is arguably the best-known result from the last twenty-five years of organizational behavior research into motivation. Nonetheless, the impact of goal-setting theory has not been felt in all areas of work or academia, and its prescriptions are not necessarily embedded in industrial systems. For example, a recent field study at a water utility company with the provocative title "A New Approach to Performance Measurement and Goal Setting" (Andrews, Carpentier et al., 2001) made no reference at all to goal-setting theory. A recent article on motivating workers in warehouses published (without a byline) in a practitioner-oriented publication of the leading academic society for Industrial Engineers also made no reference to goal-setting research in its prescriptions (IIE Solutions, 1999). We point these articles out, not as a criticism of the work (the validity of which does not reside in the author's reference to goal-setting), but to demonstrate the relatively impermeable wall which surrounds academic areas. Moreover, it is one of the points of this article that to some extent, operations management and industrial engineering models may have overlooked goal-setting prescriptions because they are difficult to model and because the goal-setting prescriptions are not as precise and simple as they seem at first glance.

There are a great number of complications to the seemingly simple idea that specific, difficult goals should improve performance. Some of these are fundamental: e.g., how "difficult" should the goal be, and by how much will performance improve? But the theory has also been extended, and many moderators (factors that may modify the basic relationship among goals, motivation and performance) have been implicated. Some of these have been researched quite deeply (e.g., whether participative goals work better than assigned goals). But others have received relatively less attention. For example, how does performance change over time in the presence of a goal that must be attained by a deadline?

One early meta-analysis (Tubbs, 1986) found that goals specifying quantity had a stronger effect (d = .845) than those specifying a time by which the task was to be completed (d = .420), indicating that the goal-setting effect is sensitive to the clock[1]. And recent work by Vancouver and his colleagues (Vancouver, Thompson et al., 2001; Vancouver, Putka et al., 2005) have examined the dynamic nature of motivation and goal-striving behavior. But, beyond the clear indication that the motivational forces in play *are* dynamic, the functional form of the relationships among goals, motivation and performance over time, especially in the presence of a deadline, is not well understood.

The implications of the lack of understanding of the impact of time as a moderator are easy to see by examining one of the moderators that (unlike time) is fairly well understood: goal difficulty. Goal difficulty has been operationalized in various ways, but a method that is comparable across studies, and, thus, useful for meta-analysis, is to use the frequency with which the goal is attained as a measure of difficulty (Wright, 1990; Wright, Hollenbeck et al., 1995). This is our parameter α. To our knowledge, there is no single prescribed number or equation for a "difficult" versus an "easy" goal in terms of α. A recent article referred to goals achieved 15% of the time or less as "difficult," and a goal achieved 50% of the time or more as "easy" (Klein, Wesson et al., 1999). But prescriptions for a "difficult" goal setting have been given as low as 10% attainment and as high as 25% attainment. Nonetheless, what is clear across many studies is that once a goal is perceived as "easy," it loses its motivational force; while once a

[1] The d statistic is a measure of effect size across studies, in terms of the number of standard deviations by which the mean impact is shifted. For an intervention with d = .52, a task with expected duration of 10 minutes and a standard deviation of 2 minutes would be expected to finish in 10 - .52(2) = 8.96 minutes.

goal is perceived as unrealistically difficult, it will simply be rejected, and again lose its motivational force.

The presence of a deadline, with tasks arriving across deadlines adds another dimension to the issue of goal difficulty. Clearly, a task arriving closer to a deadline (or one whose completion has been delayed for some reason) will be more difficult to finish than one arriving earlier. A precise level of attainment cannot be applied to any given task—some tasks must necessarily be harder than others, and tasks which start out being "easy" may end up being "difficult." Of course, one could simply set a new goal for each task as it arrives, and modify the goal over time. But given the information processing burden of this approach (as well as the cognitive burden of having to regularly check to see if your goal has changed) a method like the one discussed here, which sets a stationary goal based solely on information available before the task arrives, is preferable. The question remains, though, as to how to make such a goal motivational in aggregate for all the tasks.

In addition to time and goal difficulty, two other moderators are particularly relevant for this study. One is the difference between field applications and lab studies, with meta-analyses of field applications typically reporting smaller effect sizes. For example, Mento et al. (1987) reported an effect size of $d = .439$ in field studies but of $d = .624$ in lab studies. Likewise, Tubbs (1986) reported an effect size of $d = .520$ in the field but of $d = .897$ in the lab. Another moderator that is important here is task complexity, with simpler tasks showing a stronger effect in meta-analyses. For example, Wood et al. (1987) report performance improvements of 12.15% ($d = .76$) on simpler tasks, but only 7.8% on more complex tasks. In this particular study, effect size of $d = 0.52$ will be assumed as a realistic but conservative estimate; though the researchers recognize the exact effect will vary from setting to setting. We make this assumption without loss of generality—the procedure reported will work regardless of the anticipated effect size.

In the context of field settings, managers are especially sensitive to time-based measures that relate to customer service. In reviewing and evaluating the metrics that logistics managers use, Caplice and Sheffi (1994) said that a metric assessed the effectiveness (as opposed to merely utilization, or productivity) of a logistics operation when the metric compared output to some normative standard. Of the thirteen common effectiveness metrics they list, eight used a time period as the normative standard for comparison (e.g., orders shipped on time). The connection between the timeliness of operations and customer service is well understood in logistics and warehousing. In a comprehensive ($N = 5,531$) Confirmatory Factor Analysis undertaken as a part of the development of a survey-based scale to assess customer service for the Defense Logistics Agency, Timeliness was one of the primary factors that emerged (Mentzer, Flint et al., 1999). And Johnson and Davis (1998) describe how Hewlett-Packard tracks the timeliness of order fulfillment in logistics operations by examining not only the average on-time performance, but the entire distribution of order fulfillment times against its deadlines or customer-promised dates (which the company calls *order aging profiles*).

The NSD metric presented in the next section is a time-based effectiveness metric closely related to order aging profiles. The performance model that follows illustrates how a motivational goal on the NSD metric can change performance. As such, it begins to fill the gap between the findings of goal-setting theory and the operational metrics and policies used to monitor and control order fulfillment operations.

Performance Model

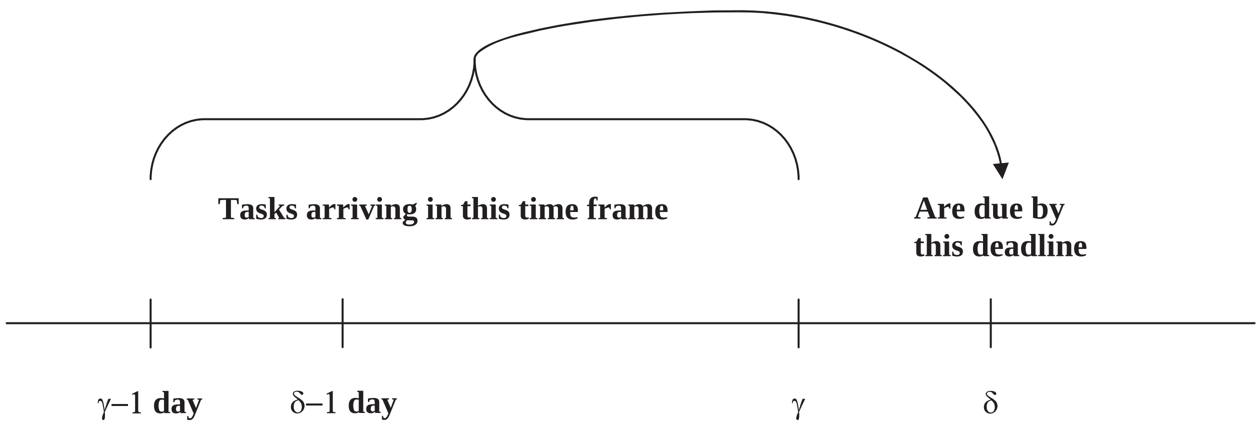
In this model, the authors will consider an order fulfillment operation to which orders arrive in a continuous stream throughout the day. Workers prepare orders and make them available for shipping. Orders ready before a specified cutoff time γ each day are batched and delivered to customers on a departing vehicle. Processing time of an order is defined as the time between its arrival and the time it is available for shipping.

For this model, γ is the cut-off time for orders due to be shipped by the departure time δ. For convenience, time is measured in days. The NSD metric N_u is the fraction of orders arriving between $\gamma-1$ and γ that are completed before δ (see Figure 1). Performance on the NSD metric for continuously arriving tasks depends on the arrival distribution $f_a(x)$ and the processing time distribution $f_p(y)$ of those tasks. The model, for now, assumes that the processing time distribution corresponds to unmotivated workers. The arrival time distribution and processing time distribution convolve to determine the completion time distribution, $f_c(x,y) = f_a(x)f_p(y)$; NSD performance is measured against this. Without accounting for motivational effects, expected performance on the NSD metric, N_u, can be expressed as

$$N_u = \iint\limits_{\substack{\gamma-1<x<\gamma \\ y<\delta-x}} f_c(x,y)\,dx\,dy \qquad \text{Equation (1)}$$

where δ is the relevant deadline. Having this expression for the *unmotivated* completion time distribution and expected NSD performance established, this study will begin to apply some of the findings from goal-setting theory to establish the cutoff time.

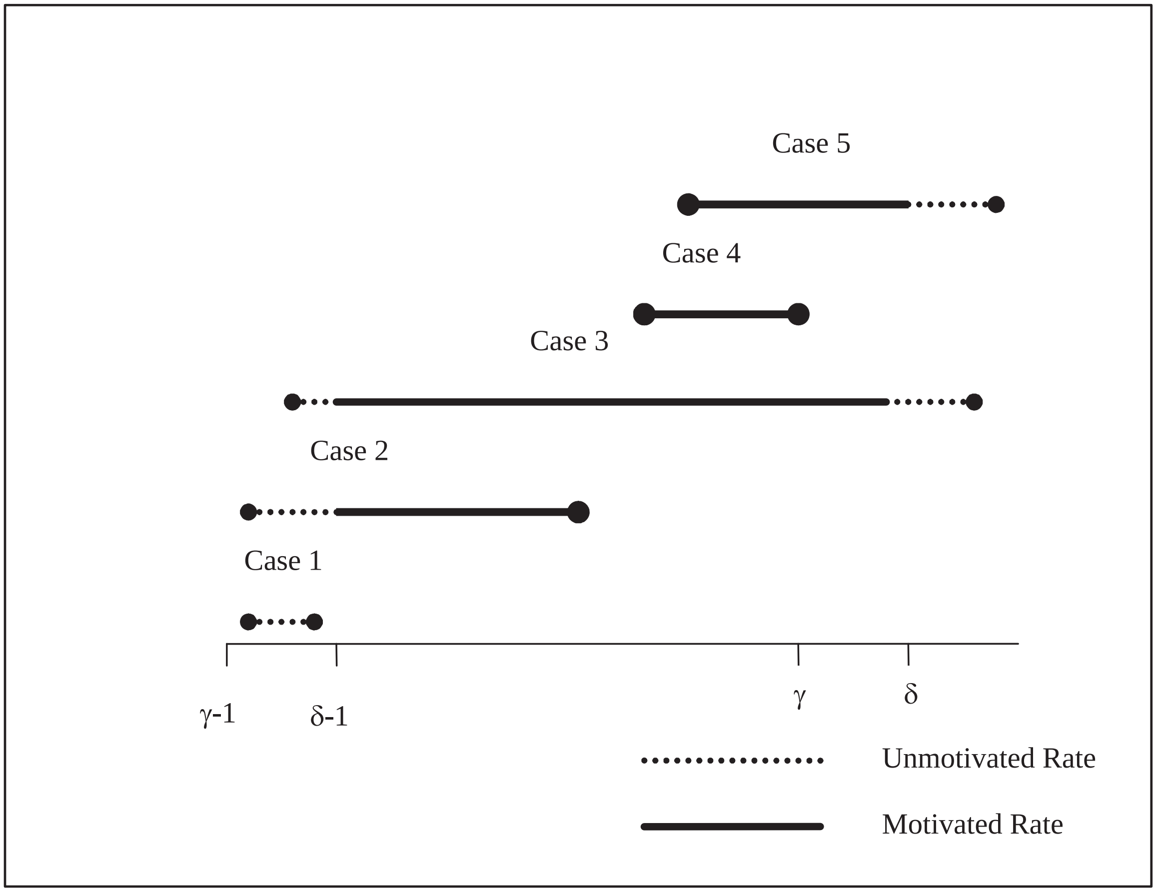
Figure 1. Cutoff times and Deadlines



As mentioned earlier, a goal is said to motivate an employee on a task in part through a mechanism which involves attention. The goal directs attention to certain tasks at the expense of (or at least to the exclusion of) other tasks. Once attention is focused on goal-relevant tasks, one can expect persistence and effort to increase and can reduce processing time on those tasks. Therefore, tasks arriving after the cutoff but before the previous deadline cannot be expected to receive the same attention as the tasks which need to be completed for the more proximal goal. Moreover, given random arrivals across the deadline periods and a relatively long processing time distribution, there is a significant chance that some tasks will not be completed by the deadline. In that case, a new goal becomes relevant, and work on the incomplete tasks can no longer help attain that new goal.

This leads to a definition of a *motivated window* of time [δ-1, δ] for tasks which arrive between [γ-1,γ]. A task that arrives between [γ-1, δ-1] will be worked on at an unmotivated rate until δ-1, because the task is not relevant to the most proximal goal.   A task which arrived before γ which is unfinished by δ will no longer be worked on at the motivated rate because it is no longer relevant for attaining the goal. Figure 2 shows the five possible combinations of motivated and unmotivated task work rates.

Figure 2. Motivated Window



The motivated rate itself can be expressed in terms of the expected motivational effect of goal setting. As explained in the previous section, this is usually expressed in terms of the meta-analysis coefficient d—that is, as a reduction in task time equivalent to d standard deviations. Given our expected unmotivated task time

$$\bar{t}_p = \int y \, f_p(y) \, dy$$

the expected motivated task time will be

$$\bar{t}_m = \bar{t}_u - d \, \sigma_p$$

where σ_p is the standard deviation of $f_p(y)$. The corresponding rates will, of course, be

$$r_u = \frac{1}{\bar{t}_p}$$ Equation (2)

and

$$r_m = \frac{1}{\bar{t}_p - d\sigma_p}$$ Equation (3)

As r_m increases, the processing time distribution $f_p(y)$ will underestimate the probability that a task will be finished within a given time period, and hence Equation (1) is an underestimate of NSD performance given the presence of a motivated window. Equations (2) and (3) are used to develop formulae for task completion time z for each of the five cases shown in Figure 3. The expressions below rely on the fact that the duration of motivated tasks (or motivated portions of tasks) is $(\frac{r_u}{r_m})$ less than the duration of unmotivated tasks. The relative probability of each of these cases occurring can be determined through the joint distribution in (1).

Case 1: $\gamma - 1 < x < \delta - 1$; $y < (\delta - 1) - x$

$$z = x + t_p$$ Equation (4)

Case 2: $\gamma - 1 < x < \delta - 1$; $(t_p - (\delta - 1) - x) (\frac{r_u}{r_m}) < 1$

$$z = \delta - 1 + (t_p - (\delta - 1) - x) (\frac{r_u}{r_m})$$ Equation (5)

Case 3: $\gamma - 1 < x < \delta - 1$; $(t_p - (\delta - 1) - x) (\frac{r_u}{r_m}) > 1$

$$z = \delta + t_p - ((\delta - 1) - x) - \frac{r_m}{r_u}$$ Equation (6)

Case 4: $x > \delta - 1$; $t_p (\frac{r_u}{r_m}) < \delta - x$

$$z = x + t_p \left(\frac{r_u}{r_m}\right) \qquad \text{Equation (7)}$$

Case 5: $x > \delta - 1;\ t_p \left(\frac{r_u}{r_m}\right) > \delta - x$

$$z = \delta + \left(t_p - \left((\delta - x)\left(\frac{r_m}{r_u}\right)\right)\right) \qquad \text{Equation (8)}$$

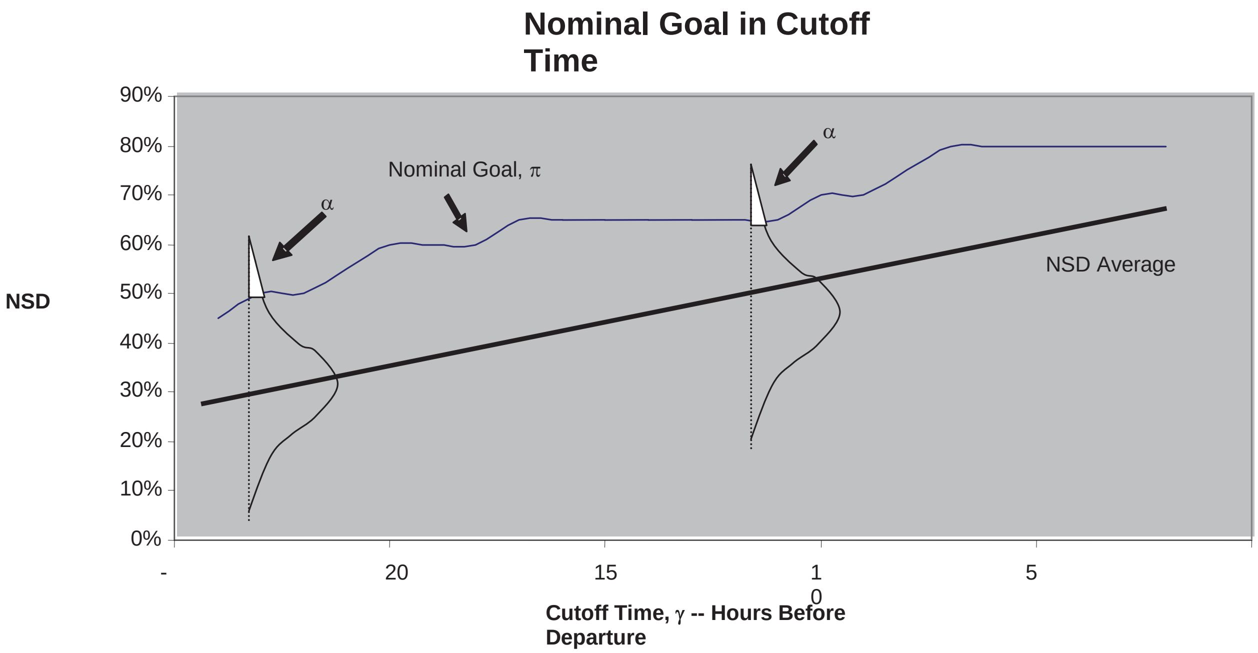
Given a cutoff time γ, these expressions can be used to predict motivated NSD performance N_m against a nominal goal π by using a Monte Carlo simulation to integrate the cases across the joint distribution of arrival times and (unmotivated) processing times. But this model has not yet dealt with the problem of how to determine values for γ or π.

THIS PAGE INTENTIONALLY LEFT BLANK

A Goal-setting Procedure

As stated in the above discussion on goal-setting literature, a motivational goal needs to be difficult, and goal difficulty is usually defined in terms of α, the frequency of goal attainment. A "difficult" goal is one that has a relatively small chance of being attained (e.g., α = 20%). However, having addressed an appropriate level of goal attainment, this model highlights more than one way to set π and γ to create a policy that has the desired level of difficulty. In other words, a given nominal goal π will be more easily attained by moving the cutoff time γ backward in time. Conversely, a given γ can be made to correspond to a relatively easier goal by lowering π. The relationships among π, γ, and α are shown in Figure 3.

Figure 3. Relationships among γ, π and α



Note: in general, the average value of N_m is not equal to π, and performance is below the nominal goal (1-α)% of the time.

Goal-setting theory prescribes not only a level for α, but a level for γ as well (see Figure 2). Clearly, performance improvement is maximized when the motivational window is as wide as possible—that is, when δ - γ is as small as possible. Viewing Figure 3, this seems counter-intuitive—setting the cutoff time close to the deadline is associated with the lowest performance against the nominal goal. However, it is important to remember that the target deadline for a given set of orders may shift as one looks across Figure 3 from left to right. In other words, Figure 3 shows performance against a deadline determined by the cutoff time, not the actual percentage of tasks accomplished by the deadline that immediately follows the arrival of the task (we explore the implications of this in the next section).

Hence, goal-setting theory prescribes values for both α and γ, and in so doing, asserts an appropriate value for π. It is the nominal goal which, given a cutoff time γ, will be achieved only about α% of the time. In Figure 4, a distribution of N_m is shown for different values of γ. On each of those distributions, π is the quantile associated with a performance that is only obtained α% of the time. The distribution of N_m is a function of the (motivated) distribution of task completion times given in equations 4-8, and π is a point on the distribution of N_m that requires estimation. As the Central Limit Theorem does not apply to quantile estimates, a procedure such as bootstrapping (Efron & Tibshirani, 1998) must be used to provide robust estimates of π.

The next section will demonstrate this procedure on a set of field data.

Application and Analysis

The authors applied the performance model and goal-setting procedure to data from a field site in California. Using the tenets of goal-setting theory, they set α to a "difficult" goal level, and made γ as late as possible in order to maximize the size of the motivated window. Given these values for α and γ, they demonstrated the use of a Monte Carlo simulation and bootstrapping procedure to estimate π and calculated the expected performance of their policy. Next, they compared their results to a policy that management might find intuitively appealing and demonstrated the superiority of their approach. Finally, they conducted sensitivity analysis to assess the effect of setting an upper bound on cutoff time or a lower bound on the magnitude of the nominal goal.

To apply the model developed in the previous section, they used the samples shown in Figures 1 and 2 as estimates of the arrival and processing time distribution functions in a Monte Carlo simulation. For simplicity, from this point forward, this discussion will state γ in terms of the hours before the deadline. To maximize the size of the motivated window, they set the cutoff time one hour before the deadline, $\gamma = 1$. They then used the simulation and equations 4-8 to derive a finish-time distribution for each task, $f_c^m(x, y)$, accounting for motivation effects. They simulated 300 tasks for each deadline, which was the average number of tasks per day at the field site, and calculated the percentage of tasks which finished by the deadline, N_m. Each run of this simulation provides this discussion with a single point on a sample distribution of N_m.

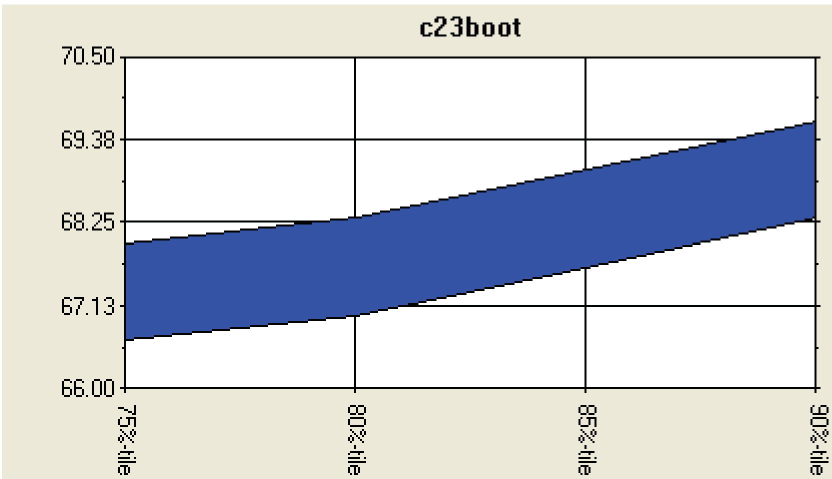
They set the desired percent-goal-attainment, α, to 20% to represent a difficult goal. Their simulation was then repeated 100 times, yielding a sampling distribution of 100 observations of performance. This study will refer to the CDF of this distribution as $\hat{\Phi}_\gamma(N_m)$, and will note that the research team was seeking $\pi = N_m \ni \hat{\Phi}_\gamma(N_m) = \alpha$. That is,

$$\hat{\Phi}_\gamma^{-1}(\alpha) = \pi \qquad \text{Equation (9)}$$

or, in this case, $\hat{\Phi}_1^{-1}(.2) = \pi$.

Note however, that $\hat{\Phi}_\gamma(N_m)$ is the sampling distribution of a point estimate, the quantile associated with the number of tasks completed by the deadline. The estimate of π is, thus, not subject to the Central Limit Theorem, and there is no guarantee that 100, or even 1000 observations would yield an efficient estimate. To assess the quality of their estimate, the field site team conducted a bootstrap analysis on their Monte Carlo simulation results, resampling $\hat{\Phi}_{\gamma}(N_m)$ 500 times to get acceptably low variance in the estimate (Efron &Tibshirani, 1998, p. 275). A 90% confidence interval is shown in Figure 4 and indicates that the nominal goal should be set at 68%. Figure 4 also shows that, on these data at least, the procedure is relatively insensitive (robust) to the particular value selected for α, which is good since (as mentioned previously) there is no strict guidance on a specific level for α.

Figure 4. Ninety-percent Confidence Intervals for π, with γ = 01, and Various Values of α



To assess the quality of the policy the researchers derived from these data ($\alpha = 0.2$, $\gamma = 1$, $\pi = 0.68$), we compared it to intuitive policies that might suggest themselves to management. To make this comparison, they needed to have a common yardstick, separate from NSD, against which to compare the policies. As one varies cutoff time and nominal goal, NSD has different meanings—comparing a policy of $\gamma = 23$, $\pi = 50\%$ (that is, up to 23 hours to get 50% of the tasks finished) with a policy of $\gamma = 1$, $\pi = 90\%$ (that is, as little as 1 hour to get 90% of the tasks finished) in terms of percent of tasks finished by the deadline clearly isn't a fair comparison. To solve this problem, we simply measured the percentage of tasks that were finished by the deadline immediately following arrival, *regardless* of whether that deadline was *targeted* for the task. This measure is related to the order aging profile metric mentioned in the literature review, and can be called tasks finished without delay[2]. The percentage of tasks finished without delay under their policy was 61.3%.

Management at the field site intuitively felt that the nominal goal π should be some high number, such as 80%, in order to be motivational. One policy (A) that was proposed was to use $f_p(y)$ to determine the maximum processing time required by at least 80% of the tasks and to set the cutoff time back from the deadline by that many hours—in this case, 21 hours before the deadline. Setting values for π and γ, of course, determines a value for α. One might think that this would yield an "easy" goal that would be attained about 50% of the time; but that isn't necessarily the case because setting the parameters in this way ignores the effect of the arrival-time distribution. However, when the research team simulated this policy with 300 tasks over 100 deadlines, they found that the goal was indeed attained 100% of the time (even without a motivational effect to reduce processing times). The minimum percentage of tasks completed by the deadline was 88.67%, and the maximum was 95.67%. Percentage of tasks finished without delay was only 41.1%, primarily due to the fact that this "easy" goal was not motivational.

The failure of Policy A suggests Policy B: to set a nominal goal of 80%, but then use Equation (9) to search over cutoff times until one finds a cutoff time that yields a difficult nominal goal of 80%. That is,

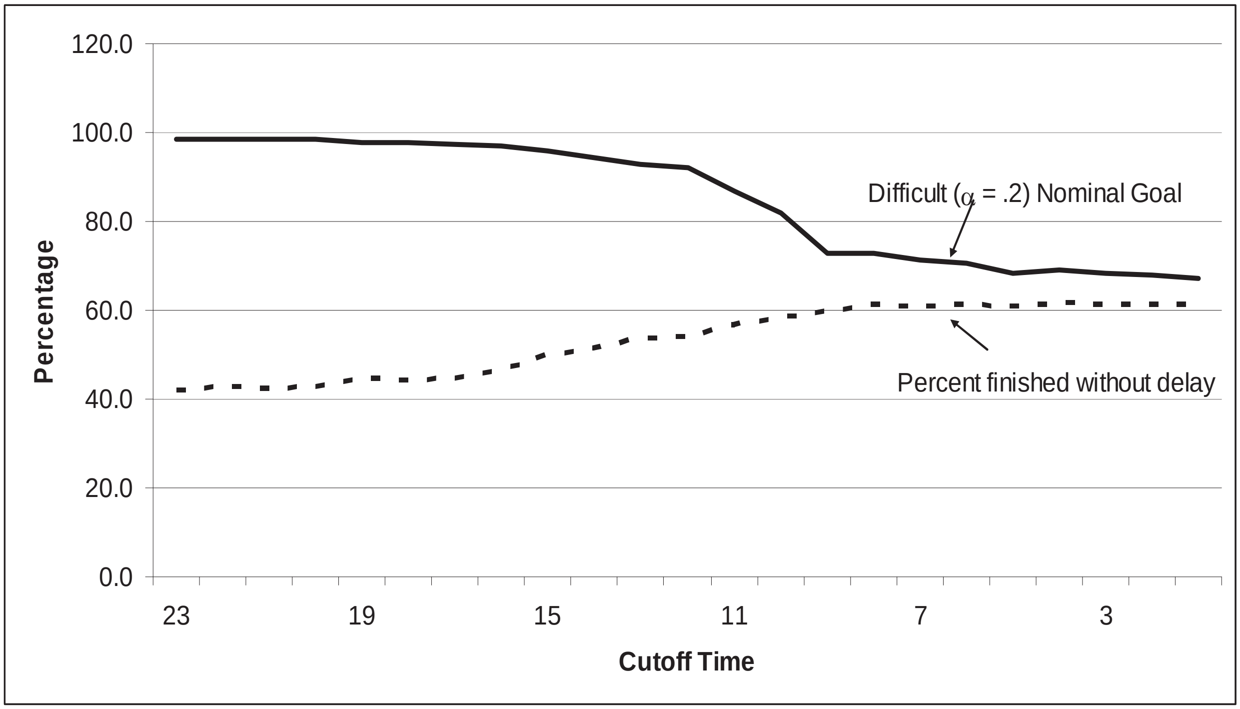
$$\gamma \ni \hat{\Phi}_{\gamma}^{-1}(.2) = 0.8 .$$

Figure 6 shows the result of this search. The cutoff time most nearly associated with a difficult goal of 80% is 10. However, the percentage of tasks finished without delay under Policy B is only 58.2%, somewhat worse than the 61.3% obtained by the team's original policy. Indeed, as Figure 5 demonstrates, the best policy is one that maximizes the size of the motivational window by setting the cutoff time as late as is practical.

---

[2] The reader might wonder why we didn't simply use this metric to begin with, set a goal against it, and avoid all the difficulty of having to determine a cutoff time. Order aging profiles are multidimensional (percent shipped over time), and it is difficult to see how one would even set a (uni-dimensional) nominal goal against them. Simply measuring tasks finished without delay is insufficient because a goal set against relatively long tasks arriving across deadlines will be rejected if the order arrives too close to the deadline—indeed, that is the whole point of setting a cutoff time. Note further that tasks finished without delay is also equivalent to NSD with the cutoff time set equal to the deadline—that is, NSD without a motivational "knob."

Figure 5. NSD Performance across Cutoff Times



In spite of the implications of Figure 5, however, there may be reasons to bound the cutoff time. Just two of these are goal acceptance and the publication of cutoff times to customers.

This study has already addressed with goal acceptance to some extent by insisting that the cutoff time be at least one hour before the deadline; we believe a goal to finish a task in one hour would most likely be rejected out-of-hand with tasks that, on average, required so much more time to complete. Goal acceptance and goal commitment have been investigated before as a moderator (Ronan, Latham et al., 1973; Klein, Wesson et al., 1999). But, beyond the general result that goals can be too difficult, and that there are individual and group differences in what constitutes too difficult a goal, there is little data specific enough upon which to build a model. This policy's one-hour limit is a recognition of the importance of goal acceptance, which requires little in the way of additional assumptions. More sophisticated approaches are possible, even given the limited data available about precisely when goals will be rejected. For example, another approach to bounding the cutoff time would be to examine the processing time distribution to find the minimum time required by at least 20% of the orders. In the field data, this was 8.1 hours. Setting the cutoff time at least 8 hours before the goal ought to give employees at least a 20% chance to finish *any* task that arrived before the cutoff time. As 20% has been shown to yield a difficult goal that is not rejected, this should provide an acceptable upper bound. As Figure 6 indicates, a cutoff time 8 hours before the deadline yields only a very small degradation in performance on the field data (tasks finished without delay drops to 59.8 from 61.3). Ultimately however, goal acceptance remains an empirical question; good scales exist, and goal acceptance is best addressed in each context by measuring it.

A more difficult question is whether or not to publish cutoff times to customers. This is done in some distribution businesses, for example, when customers are told that if they place their order by 14:00, they will receive their order the next day. In such cases, what is purely a nominal goal in our model becomes the percentage of time a customer-promised date is fulfilled; consequently, values of π as low as 68% are simply unacceptable. On the other hand, cutoff times set too early will not provide any marketing advantage (customers are likely to be unimpressed if told their order will arrive the next day, so long as they order by 1:00 a.m. on the previous day). The development of a model to capture these customer service costs is beyond the scope of the current paper, but Figure 5 makes some of the tradeoffs plain. In this researcher's data, there is very little reason to move the cutoff time earlier than 16 hours before the deadline, where the nominal goal is 97% and the percentage of tasks finished with delay is 46.5%. Cutoff times set any earlier yield minimal improvements in the nominal goal (only up to 99%) at a large expense in terms the percentage of tasks finished without delay (down to 41.9%), as well as the size of the motivational window and the "marketing opportunity" of a late cutoff time. On the other hand, after γ=12 hours before the deadline, the nominal goal drops off quite rapidly, moving from 92% down to 73% only three hours later. Between 16 and 12, the choice of a cutoff time would depend on the relative marketing value of the later cutoff time balanced against the cost of lowered delivery-as-promised from 97% to 92%. Of course, these times are wholly a function of the particular arrival and processing time distributions at the above particular field site. However, given other arrival and processing time distributions, our procedure will still yield appropriate (motivational) nominal goals for each cutoff time so that management can make a better informed decision about setting and publishing the cutoff time.

But as already pointed out, a simpler method might simply be to establish two cutoff times—there is no reason why the cutoff time given to motivate employees must be the same as the cutoff time published to customers. The intent of each is different. Employees should be aware of the other cutoff time, since the firm would expect nearly all tasks arriving before the earlier cutoff time to make the shipment. But of course, the cutoff time would have to be set earlier enough to get nearly all the orders that arrive before the customer-published cutoff time on the next shipment (i.e., the customer-published cutoff time should yield an "easy" and, hence, non-motivational goal).

Of course, the manager of an order fulfillment operation may object that there is nothing easy about meeting customer expectations for nearly perfect on-time delivery. Partly this is just confusion raised because of the way the word “easy” has been defined in the goal-setting literature—it is not “easy” (as the word is commonly understood) to consistently and reliably meet any performance target. But this issue of what is “easy” also raises a different managerial issue: the difference between a job standard (a minimum expectation for all fully trained employees) and goals (a stretch target that employees are expected *not* to attain, most of the time). Order fulfillment operations that struggle greatly to keep their promised delivery dates may be confusing these two things and publishing targets to customers that they do not have the capacity to “easily” (i.e., consistently and reliably) meet.

Finally, a significant complicating issue is how the published cutoff time would affect the arrival distribution itself. For example, shortly after one particular distributor published a cutoff time, many customers delayed their orders until just before the published time. The result, of course, was greater difficulty making good on the service promise. This issue needs further investigation, but is beyond the scope of the current work.

Conclusion

Before summarizing the contributions of this research, this discussion should review the limitations of this initial investigation. The first concerns the external validity, or generalizability of the researchers' results. Of course, external validity can never be rigorously addressed by a single study and is always accomplished through cross-validation. However, our findings, though based on field data, are based on a simulation model of the process in question and not experimentation with the process itself. While the use of simulation models to investigate behavioral phenomenon is becoming increasingly common and more widely accepted (e.g., Vancouver, Putka et al., 2005), such research will always particularly need further cross-validation from field and laboratory work. The field data contained in this study also come from a unique setting; and the particular values of the parameters the team derived (i.e., π and γ), as well as the particular performance improvement they reported, are, of course, unique to that setting as well. The unusual nature of the field setting has strengths as well as weaknesses, however. Having demonstrated that this performance model and bootstrapping procedure is robust enough to yield good results with these unusual distributions, this study suggests there is no reason to suppose that they could not be even more easily applied to a situation where tasks and arrivals followed a more predictable pattern.

Second, this study assumes particular values for the effect size d and goal attainment α. While the choice was made from within a range of reported values, the particular numbers chosen were somewhat arbitrary. Of course, if the empirical results of goal-setting research are ever to be applied in modeling work, similar assumptions will have to be made; we are not aware of any more rigorous procedure by which he could have selected a value for d or α. It might be argued that sensitivity analysis should have been conducted on these parameters. However, a quick look at the above performance model (Figure 2 and Equations 4-8) should convince the reader that although the magnitude of improvement obtained by the field site procedure might change if d were different, any positive value for d would have yielded a similar result to that found in Figure 5: the wider the motivational window, the better. Figure 4 provides at least a limited sensitivity analysis on α; but again, it should be clear that the bootstrapping procedure illustrated in Equation 9 will provide a motivational goal, regardless of the selection of α, so long as α is chosen on a level that is neither too easy, nor too hard.

Finally, while we believe the performance model developed in this study is a fair representation of the current state of knowledge of goal-setting effects in a dynamic setting, it does not use the sorts of information technology available today to modify goals based on information about particular orders or the current state of the operation. In defense of this, it should be pointed out that there is no clear guidance, currently, on what to do with such information in terms of setting goals. Important extensions are needed to goal-setting theory in general in order to investigate exactly how motivation shifts over time, especially as a deadline approaches. The model discussed in this study, in which motivation is either present or absent depending on the current time in relation to the targeted deadline, could be elaborated were such information available. While the simulation work of Vancouver et al. (2005) is an important step in showing how motivational changes over time might impact performance, empirical work is needed to assess the exact functional form of the changing levels of motivation over time when a goal is given against a deadline.

In summary, an order fulfillment system with order deadlines must have a sensible performance metric for customer service. In order to be effective, this metric must incorporate the linkage between upstream continuous processes and downstream batch processes. The Next Scheduled Deadline metric defined in this study incorporates these processes in a way that allows managers to adjust to any order arrival or processing time distribution. When published to workers (along with a properly established goal), the metric encourages motivated behaviors such as increased work rates and improved task strategies. This motivation, in turn, improves customer service by causing more customers to receive their orders sooner. An important feature of the metric is that, for a given cutoff time, an increase in its recorded value necessarily means an improvement in real customer service—in this case, reduced customer waiting time.

This procedure takes as input the distributions of task arrival and processing times. Using principles from goal-setting theory, the researchers developed a performance model (Equations 4-8) and used it to set a cutoff time between deadlines. Tasks arriving after the cutoff time are not targeted for the immediately following deadline, but the subsequent one. To set a motivational goal for each deadline, we again applied principals from goal-setting theory to determine a desired level of difficulty, then applied a bootstrapping procedure (Equation 9) to estimate a nominal goal corresponding to that level of difficulty for the given cutoff time.

We compared these results to two policies that might have intuitive appeal to management based on the percentage of tasks finished without delay (i.e., by the immediately following deadline, regardless of which deadline was “targeted”’) and found that this procedure was superior. Under this procedure, 61.3% of tasks were finished without delay—compared to 41.1% and 58.2% under the intuitive policies.

References

Andrews, B. H., Carpentier, J. J., et al. (2001). A new approach to performance measurement and goal setting. *Interfaces*, *31*(3), 44-54.

Austin, R. & Larkey, P. (2002). The future of performance measurement: Measuring knowledge work. In A. Neely (Ed.), *Business Performance Measurement*. Cambridge, MA: University of Cambridge.

Boudreau, J., Hopp, W. J., et al. (2003). On the interface between operations and human resources management. *Manufacturing & Service Operations Management 5*(3), 179-202.

Caplice, C. & Sheffi,Y. (1994). A review and evaluation of logistics metrics. *The International Journalof Logistics Management 5*(2), 11-28.

Editors (1999). How to motivate warehouse workers. *IIE Solutions 31*(6), 9.

Efron, B. & Tibshirani, R. J. (1998). *An introduction to the bootstrap.* Boca Raton: Chapman & Hall CRC.

Euske, K. J. & Zander, L. A. (In press). History of performance measurement. In K. Kempf-Leonard (Ed). *Encyclopedia of Social Measurement.* Amsterdam: Elsevier.

Johnson, D. W., Maruyama, G., et al. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin 89*(1), 47-62.

Johnson, H. T. & Kaplan, R. S. (1987). *Relevance lost: The rise and fall of management accounting*. Boston, MA: Harvard Business School Press.

Johnson, M. E. & Davis, T. (1998). Improving supply chain performance by using order fulfillment metrics. *National Productivity Review* (1998, Summer), 3-16.

Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal* 18(4), 769-783.

Kerr, S. & Landauer, S. (2005). Using stretch goals to promote organizational effectiveness and personal growth. *Academy of Management Executive 18*(4), 134-139.

Klein, H. J., Wesson, M. J., et al. (1999). Goal commitment and the goal-setting process: Conceptual clarification and empirical synthesis. *Journal of Applied Psychology* 84(6), 885-896.

Locke, E. A., Latham, G. P., et al. (1990). A theory of goal setting and task performance. New York: Prentice Hall.

Mentzer, J. T., Flint, D. J., et al. (1999). Developing a logistics service quality scale. *Journal of Business Logistics 20* (1), 9-32.

Meyer, C. (1998). How the right measures help teams excel. *Measuring Corporate Performance*. Boston: Harvard Business School Press.

Mitchell, T. R. & Daniels, D. (2001). Motivation. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.).*Industrial and Organizational Psychology* (p. 12). New York: John Wiley & Sons, Inc.

Moinzadeh, K., Klastorin, T. D., et al. (1997). The impact of small lot ordering on traffic congestion in a physical distribution system. *IIE Transactions 29*(8), 671-680.

Neely, A. & Austin, R. (2002). Measuring performance: The operations perspective. In A. Neely (Ed.) *Business Performance Measurement*. Cambridge, MA: Cambridge University Press.

Ronan, W. W., Latham, G. P., et al. (1973). Effects of goal setting and supervision on worker behavior in an industrial setting. *Journal of Applied Psychology 58*, 302-307.

Stalk, G. & Hout, T. (1990). *Competing against time*. New York: Free Press.

Tubbs, M. E. (1986). Goal-setting: A meta-analytic examination of the empirical evidence. *Journal of Applied Psychology* 71(3), 474-483.

Vancouver, J. B., Putka, D. J., et al. (2005). Testing a computational model of the goal-level effect: An example of a neglected methodology. *Organizational Research Methods 8*(1), 100-127.

Vancouver, J. B., Thompson, C. M., et al. (2001). The changing signs in the relationships between self-efficacy, personal goals and performance. *Journal of Applied Psychology 86*, 605-620.

Wood, R. E., Locke, E. A., et al. (1987). Task complexity as a moderator of goal effects: A meta-analysis. *Journal of Applied Psychology 72*(3), 416-440.

Wright, P. M. (1990). Operationalization of goal difficulty as a moderator of the goal difficulty-performance relationship. *Journal of Applied Psychology* 75(3), 227-234.

Wright, P. M., Hollenbeck, J. R., et al. (1995). The effects of varying goal difficulty operationalizations on goal setting outcomes and processes. *Organizational Behavior and Human Decision Processes 61*(1), 28-43.

Initial Distribution List

1.	Defense Technical Information Center 8725 John J. Kingman Rd., STE 0944; Ft. Belvoir, VA 22060-6218	2
2.	Dudley Knox Library, Code 013 Naval Postgraduate School, Monterey, CA 93943-5100	2
3.	Research Office, Code 09 Naval Postgraduate School, Monterey, CA 93943-5138	1
4.	Robert N. Beck Dean, GSBPP 555 Dyer Road, Naval Postgraduate School, Monterey, CA 93943-5000	1
5.	Doug Brook Professor, GB/ 555 Dyer Road, Naval Postgraduate School, Monterey, CA 93943-5000	1
6.	Bill Gates Associate Dean for Research, GB/Gt 555 Dyer Road, Naval Postgraduate School, Monterey, CA 93943-5000	1
7.	Kenneth H. Doerr Associate Professor, GB/ 555 Dyer Road, Naval Postgraduate School, Monterey, CA 93943-5000	1
8.	Dr. Kevin Gue Associate Professor, Department of Industrial and Systems Engineering Auburn University	1
9.	Karey L. Shaffer Program Manager, GB/Ks 555 Dyer Road, Naval Postgraduate School, Monterey, CA 93943-5000	1

Copies of the Center for Defense Management Reform Research Reports may be printed from our website www.nps.navy.mil/gsbpp/CDMR

THIS PAGE INTENTIONALLY LEFT BLANK

2006 Research Products of the Center for Defense Management Reform

Published Student Research

NPS-CDMR-GM-06-004	*Center for Navy Business Excellence: A Catalyst for Business Transformation.* Lieutenant Gordon E. Meek, US Navy, December 2005.
NPS-CDMR-FM-06-003	*Federal Financial Reform. Policy Formulation to Implementation: Research into Relationships between the President's Management Agenda Scorecard, Federal Audited Financial Statements, and GAO High Risk List.* Captain Andrew Lind, US Air Force, December 2005.
NPS-CDMR-HR-06-002	*Performance Based Pay for the U.S Marine Corps.* Major Henry Brown, US Marine Corps, Captain Owen Nucci, US Marine Corps, December 2005.
NPS-CDMR-HR-06-001	*The Department of Homeland Security: A Gateway to Civil Service Reform.* Lieutenant David W. Anderson, US Navy, Captain Joshua P. Bahr, US Marine Corps, December 2005.

Technical Reports

NPS-CDMR-LM-06-005	*A Performance Metric and Goal Setting Procedure For Deadline-Oriented Processes*. Kenneth H. Doerr, PhD, Associate Professor, Kevin R. Gue, PhD, Associate Professor, December 2005.

Copies of the Center for Defense Management Reform Research Reports may be printed from our website www.nps.navy.mil/gsbpp/CDMR

THIS PAGE INTENTIONALLY LEFT BLANK

CENTER FOR DEFENSE MANAGEMENT REFORM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CALIFORNIA 93943

www.nps.navy.mil/gsbpp/cdmr